

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Modélisation de la mobilité résidentielle en vue d'une micro-simulation des évolutions de population

Cornelis, Eric; Barthelemy, Johan; Pauly, Xavier; Walle, Fabien

*Published in:*  
les cahiers scientifiques du transport

*Publication date:*  
2012

*Document Version*  
le PDF de l'éditeur

[Link to publication](#)

*Citation for pulished version (HARVARD):*

Cornelis, E, Barthelemy, J, Pauly, X & Walle, F 2012, 'Modélisation de la mobilité résidentielle en vue d'une micro-simulation des évolutions de population', *les cahiers scientifiques du transport*, VOL. 62, p. 65-84.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## **MODÉLISATION DE LA MOBILITÉ RÉSIDENTIELLE EN VUE D'UNE MICRO-SIMULATION DES ÉVOLUTIONS DE POPULATION**

**ERIC CORNELIS, JOHAN BARTHELEMY,  
XAVIER PAULY, FABIEN WALLE**

**NA~~X~~YS**

**GRT – UNIVERSITÉ DE NAMUR (FUNDP)**

### **1. INTRODUCTION**

Dans le cadre du projet MOBLOC (TOINT et alii, 2011) financé par la Politique scientifique fédérale belge (BELSPO), l'objectif poursuivi consistait à modéliser non seulement la mobilité résidentielle et la mobilité quotidienne, le transport mais également les interactions entre ces deux types de mobilité. Cette modélisation concerne l'ensemble du territoire belge et est réalisée sur un maillage géographique fin, à savoir les 589 communes du Royaume de Belgique. Pour ce faire le GRT était associé avec des démographes du GédAP de l'UCL et des géographes du CEPS-INSTEAD. Une fois un modèle conjoint construit, établissant les impacts croisés des mobilités résidentielles et quotidiennes, il faut pouvoir l'utiliser pour tester les

effets de certaines politiques et établir des prospectives à plus long terme. Si les modèles ont pu être calibrés sur des données existantes, il faut disposer d'informations sur l'évolution, notamment et surtout de la population et de ses caractéristiques retenues comme facteurs déterminants dans la modélisation, afin d'être capable d'appliquer notre outil à des projections. Cela demande également une vue spatialement désagrégée sur la population (de la Belgique) puisque tant les choix résidentiels que les besoins de mobilité sont intimement liés à l'espace. Dans notre cas, et vu les données disponibles, notre maillage géographique est celui des communes belges (NUTS4) qui sont au nombre de 589. Étant donné qu'il est impossible de disposer de données exhaustives rencontrant toutes ces contraintes et que d'ailleurs, même si ces données existaient, les lois concernant le respect de la vie privée ne nous permettraient pas de les exploiter, il nous a fallu recourir à une méthodologie permettant de construire une population synthétique (encore appelée virtuelle) de la Belgique. Cette population synthétique devait être construite avec le niveau de désagrégation spatiale souhaité, mais aussi avec toutes les caractéristiques nécessaires à l'application de notre modèle. Il fallait également mettre au point les processus permettant de faire évoluer cette population dans le temps afin de l'utiliser dans des exercices de prospective. Le développement de ces méthodologies a bénéficié des efforts de recherches accomplis dans le cadre du projet DIDAM (BARTHELEMY et alii, 2010) financé par une action de recherche concertée (ARC) réunissant le GRT (FUNDP) et le GTM (FUCAM).

Notre approche se place dans la lignée des modèles de micro-simulation qui abordent des réalités complexes, comme peuvent l'être le transport ou la mobilité résidentielle en partant du niveau individuel (ORCUTT, 1957). Elle se veut aussi dynamique en prenant en compte les évolutions temporelles des individus et des ménages et les impacts sur les phénomènes étudiés. Elle se place ainsi, très modestement, dans la lignée de modèles débutée par DYNASIM (ORCUTT et alii, 1976) et dans le courant de micro-simulation en sciences sociales tel qu'il est décrit dans SPIELAUER (2009). L'originalité de notre contribution consiste d'une part dans le développement et l'utilisation d'une méthode novatrice pour générer une population synthétique et, d'autre part, dans le fait que nous modélisons les phénomènes de mobilité résidentielle sur l'ensemble de la Belgique, territoire sur lequel cette question n'avait pas encore été étudiée avec une telle approche de micro-simulation.

Dans cet article, nous allons tout d'abord présenter la méthodologie développée par le GRT pour construire une population synthétique de la Belgique. Dans cette partie, nous justifierons aussi pourquoi une nouvelle méthode a dû être mise au point et pourquoi les techniques habituellement utilisées n'étaient pas pertinentes dans notre cas. Ensuite nous esquisserons les procédés qui doivent maintenant être mis en œuvre pour faire évoluer dans le temps cette population synthétique. Dans une seconde partie, nous nous

attacherons alors plus spécifiquement à la mobilité résidentielle ; après un rapide survol du schéma général du projet MOBLOC, nous nous attacherons à décrire le modèle de mobilité résidentielle mis en place en détaillant ses deux composantes : la propension à migrer et la localisation. Enfin, nous conclurons par quelques perspectives à la fois pour de futures améliorations des modèles mais aussi pour de possibles applications des modèles développés.

## 2. POPULATION SYNTHÉTIQUE

### 2.1. MOTIVATION

Un exercice de micro-simulation consiste tout d'abord à construire un jeu de données reprenant les caractéristiques pertinentes des agents considérés et ensuite à simuler le comportement de ces agents et à mettre à jour leurs caractéristiques en fonction des effets de leurs actions. Dans le cas qui nous préoccupe ici, les mobilités tant résidentielles que quotidiennes, ce sont les individus et/ou les ménages qui constituent les agents que devra mettre en œuvre la simulation. Indépendamment d'autres caractéristiques dont les impacts sur ces deux types de mobilité devront être déduits d'observations ou des modèles qui seront construits, il est clair que la dimension spatiale est cruciale dans ce type de simulation. En effet les mobilités sont en quelque sorte une manière d'interagir avec l'espace, d'utiliser celui-ci. Puisqu'il s'agit ici de micro-simulation, ces interactions devront nécessairement être considérées à un niveau de désagrégation spatiale assez fin. Dans le cas contraire, les variations de comportements des agents ne pourraient pas être appréhendées car l'agrégation spatiale masquerait toute modification locale. Afin de rencontrer cet objectif, nous travaillerons donc sur un maillage géographique correspondant aux 589 communes belges (NUTS4). Aller plus loin encore dans la désagrégation spatiale ne nous semble pas pertinent notamment parce qu'il est extrêmement rare de pouvoir disposer de données avec une granularité plus fine que la commune.

Nous aurons donc besoin pour notre micro-simulation d'une population d'individus, de ménages caractérisés par une série d'attributs et localisés au niveau de la commune. L'idéal serait bien entendu de pouvoir disposer d'une base de données exhaustive reprenant toutes les caractéristiques nécessaires de la population. Il n'en est bien entendu rien. Mais même si une telle base existait, son utilisation n'en demeurerait pas moins problématique ; en effet, le respect des législations relatives à la vie privée ne permettrait vraisemblablement pas d'employer ces informations. Et si des groupes bien particuliers pouvaient accéder, moyennant des accords stricts, à ce type de bases de données, cet accès ne serait certainement pas accordé à l'ensemble de la communauté scientifique, à l'ensemble des acteurs devant baser leurs études sur ce type d'informations.

Il faut donc recourir à une méthodologie permettant de pallier ces inconvénients. Celle-ci devra nous offrir la possibilité de travailler avec un portrait approché de la population aussi proche de possible de la réalité. Construire une population synthétique, ou virtuelle, est un outil visant à rencontrer cette attente (CIRILLO et alii, 2012). Nous allons maintenant brièvement présenter quelques approches classiques de cette méthode avant de voir pourquoi celles-ci ne sont pas pertinentes dans notre cas pour, ensuite, décrire la nouvelle méthodologie développée par le GRT et son originalité.

## 2.2. APPROCHES CLASSIQUES

La plupart des constructions de populations synthétiques se basent sur les principes de l'IPFP (*Iterative Proportion Fitting Process*) issus des travaux de DEMING et STEPHAN (1940) comme on peut le voir dans BECKMAN et alii (1996), WILSON et POWNALL (1976) ou FRICK et AXHAUSEN (2004). En gros, il s'agit de disposer d'un échantillon représentatif de la population et des totaux marginaux correspondant aux attributs caractérisant la population. On tire alors aléatoirement un individu dans l'échantillon représentatif (tirage avec remplacement) et on le place dans la catégorie de la population à laquelle il appartient; une catégorie correspondant au croisement d'une modalité de chacun des attributs pris en considération. Cette procédure est répétée jusqu'à ce que la population formée des individus ainsi tirés soit telle que les totaux marginaux connus soient égaux à ceux calculés pour la population réelle. En caricaturant, on pourrait dire qu'il s'agit d'une méthode de construction d'une population par clonage des individus d'un échantillon. Il découle immédiatement de cette technique que si des catégories ne sont pas présentes dans l'échantillon, elles ne le seront pas non plus dans la population construite.

Cette méthode « classique » ne permet de construire qu'une population d'individus ou de ménages mais pas une population « à deux niveaux » avec des individus rassemblés dans des ménages. Pour pallier cet inconvénient, GUO et BHAT (2007) ont proposé une nouvelle technique, à nouveau basée sur l'IPFP mais permettant de construire à la fois des individus et des ménages synthétiques.

D'autres méthodes se basent, elles, sur de l'optimisation combinatoire (VOAS, WILLIAMSON, 2001 ; HUANG, WILLIAMSON, 2002) mais requièrent également un échantillon de la population.

## 2.3. PROBLÈMES POUR L'APPLICATION DES APPROCHES CLASSIQUES AU CAS DE LA BELGIQUE

Si des inconvénients intrinsèques au recours à l'IPFP, principalement l'absence, dans la population construite, de catégories non observées dans l'échantillon utilisé, ont déjà été évoqués, il n'en demeure pas moins vrai que, même si ces approches classiques étaient la panacée, elles ne pourraient

être employées dans le contexte belge. En effet, elles se basent toutes sur l'existence d'un échantillon représentatif de la population. Or un tel échantillon n'est pas disponible pour l'ensemble de toutes les communes belges, mettant ainsi à mal l'utilisation de ces méthodes.

Par ailleurs, celles-ci nécessitent également des totaux marginaux cohérents entre eux. L'expérience nous a montré que les données disponibles pour la Belgique proviennent de sources différentes, ont été récoltées à des périodes différentes et de ce fait peuvent présenter des incohérences au niveau des totaux marginaux qui peuvent en être tirés. Des exemples de telles incohérences peuvent être trouvés dans (CORNELIS et alii, 2005).

#### 2.4. NOUVELLE MÉTHODE DE CONSTRUCTION D'UNE POPULATION SYNTHÉTIQUE

Tout cela nous a amenés à développer une méthode originale de construction d'une population synthétique qui ne souffrirait pas des problèmes évoqués pour les approches classiques, notamment dans le cas de la Belgique. Par conséquent, cette méthodologie doit

- . ne pas avoir besoin d'un échantillon représentatif de la population ;
- . être applicable dans des situations où des totaux marginaux peuvent présenter des incohérences les uns par rapport aux autres ;
- . permettre de construire une population synthétique d'individus regroupés en ménages en tenant compte d'informations (de totaux marginaux) se rapportant, pour certains à l'un de ces niveaux (les individus), pour d'autres, à l'autre niveau (les ménages).

Cette nouvelle technique développée par le GRT est décrite en détails dans (BARTHELEMY, TOINT, 2012) ; nous nous contenterons donc ici d'en présenter les principales étapes.

Le principe sous-jacent au modèle est très simple : il s'agit de générer les individus et les ménages en tirant leurs caractéristiques aléatoirement dans les distributions pertinentes et ce au niveau de désagrégation spatiale le plus fin disponible (c.-à-d. que si certains totaux marginaux sont disponibles au niveau communal, d'autres, par contre, ne sont accessibles qu'au niveau des arrondissements -NUTS3) ; il convient donc de pouvoir travailler à différents niveaux de désagrégation spatiale en supposant, quand l'information n'est pas disponible au niveau le plus fin, que les distributions connues à un niveau plus agrégé s'appliquent uniformément en chacune des entités de « niveau inférieur ». Lors de cette génération, il convient également de maintenir les structures de corrélation connues, c.-à-d. de respecter les totaux marginaux croisés qui ont été mesurés.

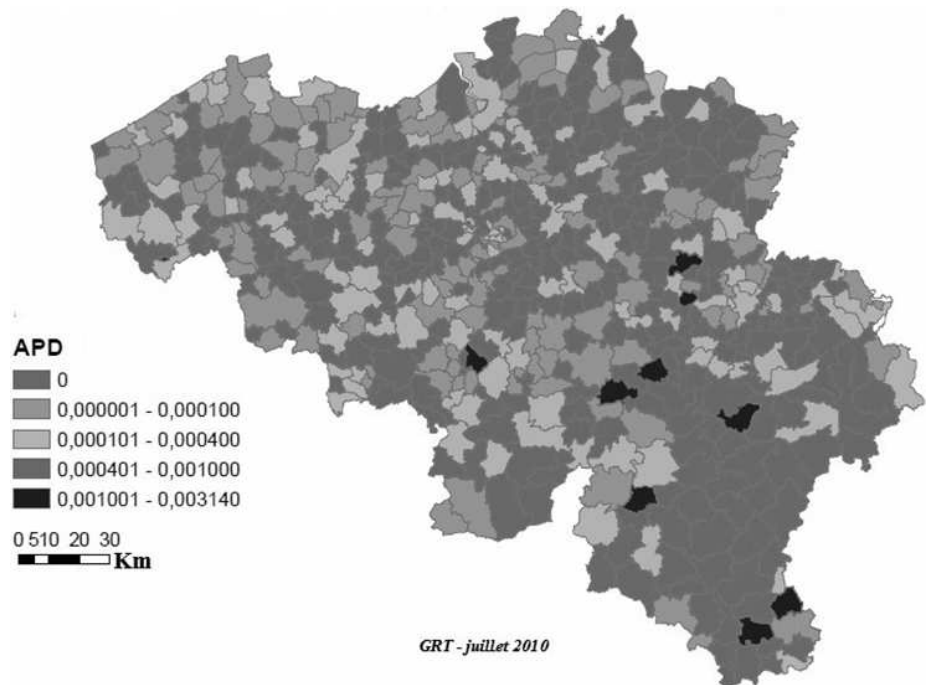
Concrètement, la procédure développée par le GRT s'articule autour de trois étapes entreprises consécutivement pour chacune des 589 communes :

- a. la génération d'un pool d'individus en respectant les totaux marginaux connus au niveau des individus (soit au niveau communal soit au

- niveau des arrondissements);
- b. l'estimation des distributions croisées des ménages ;
- c. la construction des ménages synthétiques en tirant leurs membres dans les individus du pool.

La Figure 1 nous montre les différences absolues en pourcentage (APD) [c.-à-d. la valeur absolue du rapport de la différence entre la valeur estimée (x) et la valeur réelle (y) sur la valeur estimée) pour le nombre d'individus dans chacune des communes belges. Plus précisément ce sont les différences entre les tables de contingence estimées dans les étapes a. et b. et celles résultant de l'étape de génération des ménages (c.) qui sont mesurées. Les différences observées (entre la population « réelle » et la population synthétique) sont toutes inférieures à un demi pour-cent. Dans la majorité des cas, nous sommes même en deçà du dixième de pour-cent. D'autres tests statistiques, ainsi qu'une comparaison avec les méthodes IPF, ont permis de vérifier la validité de la méthode ; ces validations sont reprises dans BARTHELEMY, TOINT (2012).

*Figure 1 : Différences absolues en pourcentages entre population « réelle » et population synthétique en considérant le nombre estimé d'individus et le nombre d'individus générés*



## 2.5. ÉVOLUTIONS DE LA POPULATION SYNTHÉTIQUE

La méthode que nous venons de décrire nous permet de créer une population synthétique pour une année de référence correspondant aux observations,

aux données que nous avons pu collecter. Cette population synthétique est donc construite et validée par rapport à des totaux marginaux que nous supposons associés à une année de base (dans la réalité, nous extrapolons un peu puisque les informations, les totaux marginaux disponibles peuvent correspondre à différentes années). Si nous voulons utiliser ce portrait, cette approximation de la population réelle dans des exercices de prospectives, de prédictions pour le futur, il convient de mettre en place des mécanismes permettant de faire évoluer temporellement la population synthétique de base.

Dans un premier temps, nous utilisons les taux de fécondité et de mortalité connus (au niveau communal) « instantanément » (c.-à-d. ceux mesurés pour l'année de référence) ou sous forme de tendances (quelle est la courbe d'évolution de ces taux sur les dernières années) afin d'estimer comment les pyramides des âges vont évoluer au cours du temps.

Pour ce qui est des autres attributs caractérisant la population (que ce soit au niveau des individus ou des ménages), différentes techniques peuvent être mises en œuvre pour modéliser leur évolution au cours du temps. C'est principalement la disponibilité des données qui amènera à recourir à l'une ou l'autre méthode d'évolution.

Cette modélisation de l'évolution temporelle de la population synthétique est en cours et nous pensons principalement à deux mécanismes pour simuler les variations de la distribution des attributs, au niveau des individus ou des ménages, dans la population synthétique :

- . les matrices de transition ;
- . la théorie des choix discrets (TRAIN, 2003).

Dans le premier cas, il s'agit, sur base d'observations, de mesurer quelles sont les probabilités de passer d'un état à un autre (p.ex., pour un ménage, de passer de l'état « sans enfant » à celui « avec un enfant ») et ce soit uniformément au sein de toute la population ou, si les données disponibles le permettent, en fonction d'autres caractéristiques (p.ex., pour l'exemple évoqué plus haut, ces probabilités pourraient être différentes en fonction de l'âge du chef de famille). Ensuite, on applique ces probabilités à la population synthétique de base pour faire évoluer celle-ci au cours du temps. Agir ainsi nécessite de poser le postulat que les tendances dans les changements d'état resteront constantes au cours du temps. Mais il s'agit là d'un scénario « au fil de l'eau » qui peut laisser place à d'autres scénarios de prospective incluant des changements de tendance dans l'avenir.

Dans le second cas, la valeur de certains attributs est estimée en fonction de celles d'autres attributs (p.ex. la possession du permis de conduire en fonction de l'âge, du genre et du diplôme). Pour ce faire, un modèle de choix discrets est calibré sur les observations dont on dispose, puis est appliqué dans les projections après avoir fait évoluer les variables explicatives suivant d'autres mécanismes (selon un scénario au fil de l'eau ou un autre).



### 3. MOBILITÉ RÉSIDENTIELLE

Les populations synthétiques, et leurs évolutions, décrites dans la section précédente sont un outil qui va nous permettre d'utiliser les modèles de mobilité résidentielle, que nous allons maintenant détailler, en vue de réaliser des prévisions sur la population belge.

#### 3.1. MOTIVATION

Les questions qui nous intéressent ici sont : quels sont les facteurs qui poussent un ménage à déménager ? Quels sont les critères pris en compte dans le choix d'une nouvelle localité de résidence ? Des études antérieures ont déjà analysé la propension à changer de résidence en fonction d'un certain nombre de caractéristiques individuelles telles que l'emploi et l'exclusion sociale (DUJARDIN et alii, 2008), les valeurs et les croyances (LINDBERG et alii, 1992), le cycle de vie (ROSSI, 1955 ; NIJKAMP et alii, 1993), la santé ou les revenus (ZEGRAS, SRINIVASAN, 2007) ou liées à l'environnement telles que la présence d'espaces ouverts ou d'aménités (CAVAILHES et alii, 2004 ; TURNER, 2005 ; CARUSO et alii, 2007 ; NG, 2008) ou la pendularité (CHOOCHARUKUL et alii, 2008) mais les effets des tendances à long terme telles que le vieillissement de la population, l'évolution de la structure des ménages/des familles sur les choix résidentiels restent à ce stade largement imprévisibles. Le projet MOBLOC a été une opportunité pour élargir la connaissance de ces phénomènes au niveau de la Belgique. Son originalité a été, entre autres, de pouvoir prendre en compte à la fois des données individuelles, d'autres relatives aux transitions dans les ménages et enfin des informations sur le logement.

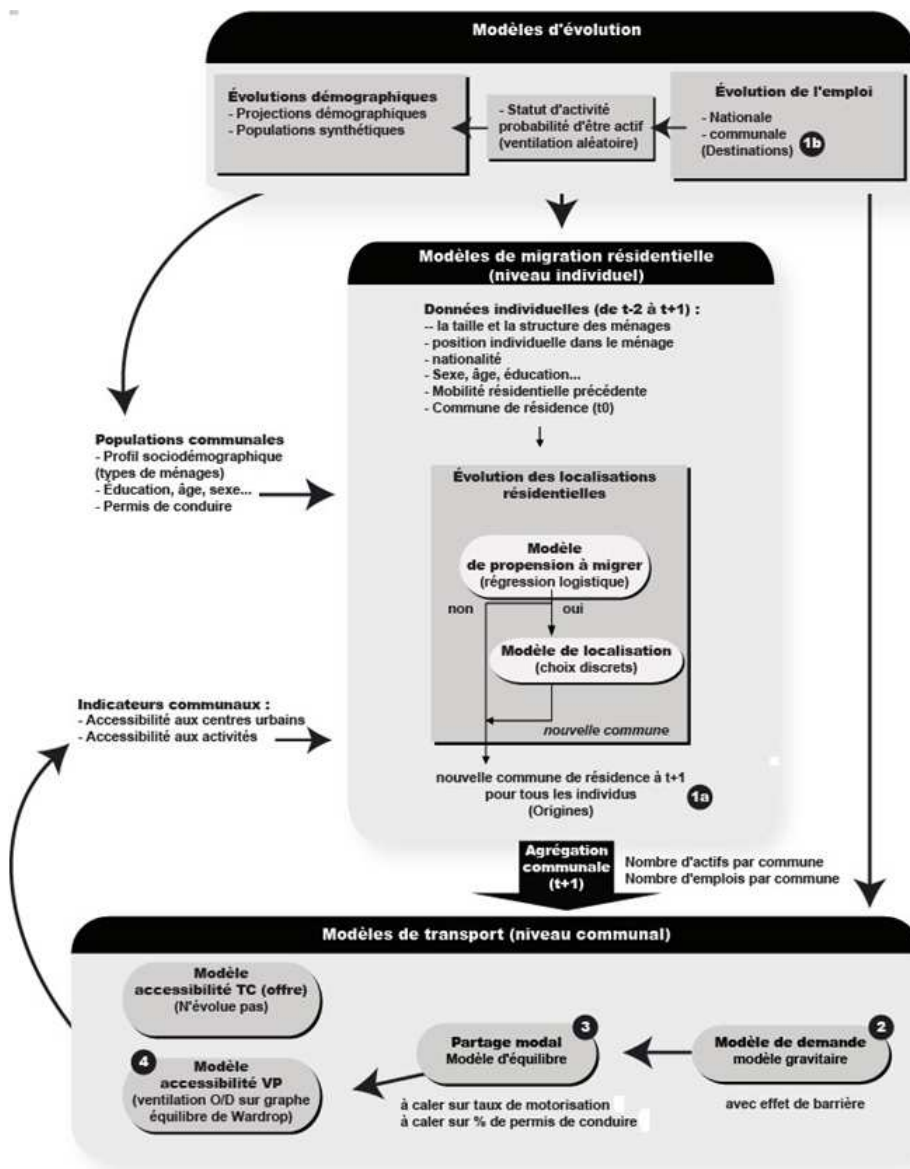
#### 3.2. MÉTHODOLOGIE GÉNÉRALE

Dans un premier temps, nous allons donner une vue d'ensemble sur la méthodologie suivie dans ce projet MOBLOC et plus précisément sur l'enchaînement des modèles et les flux des données, en entrées et en sorties, entre ces modèles. La Figure 2 résume la chaîne des modèles mis en œuvre.

Un des objectifs principaux de la recherche étant d'établir les liens entre migration résidentielle et mobilité quotidienne, les deux principales briques constituant la méthodologie sont un modèle de migration résidentielle (déménagement) et un modèle d'accessibilité. Néanmoins d'autres modèles doivent également être mis en œuvre pour déterminer certaines entrées de ces deux modèles principaux ou pour établir des liens entre ces données.

Il faut noter que le présent article se concentre sur le modèle de migration résidentielle ; nous renvoyons à KLEIN et alii (2010) pour une description plus détaillée du modèle d'accessibilité, c.-à-d. du modèle lié aux mobilités quotidiennes.

Figure 2 : Schéma général de la méthodologie suivie et des flux de données



En entrée du modèle de migration résidentielle (1a) on trouve des données individuelles. Celles-ci reprennent par exemple l'âge, le sexe, le niveau d'éducation, l'évolution du ménage, le déménagement précédent et la commune de résidence actuelle (en l'année A). Pour calibrer le modèle, des données issues des recensements de 1991 et 2001 ont été utilisées ainsi que des données provenant de la situation du Registre National à des 1er janvier d'années successives (2001-2005). Celles-ci donnent l'occasion de suivre ainsi, d'année en année, la trajectoire des individus, des ménages. Il faut

cependant noter que l'emploi de ces données ne permet de prendre en compte les déménagements que si ceux-ci sont intercommunaux (on change de commune lorsqu'on déménage) ; on néglige donc les changements de domicile à l'intérieur d'une même commune. Par conséquent, dans notre modélisation, et donc dans la suite de cet article, il faut entendre par mobilité résidentielle, un changement de domicile impliquant également un changement de commune.

Ce modèle fournit comme résultat la (nouvelle) commune de résidence (un an plus tard, en l'année  $A+1$ ). Plus précisément, ce modèle de mobilité résidentielle est constitué de deux sous-modèles : un modèle de propension à migrer et un modèle de localisation.

Le premier se sert d'un certain nombre des données individuelles pour prédire si un déménagement (c.-à-d. si un individu va changer de commune de résidence) va survenir entre l'année  $A$  et l'année  $A+1$ . Ce modèle est basé sur une régression logistique binaire. Il incorpore entre autres une composante dynamique en tenant compte non seulement du passé (déménagement précédent, modification dans la structure familiale -mariage, séparation, etc.) mais également d'événements anticipés (p.ex. une naissance prévue).

Si le modèle de propension à migrer indique qu'un déménagement va avoir lieu, le modèle de localisation simule alors le choix de résidence (plus exactement de commune de résidence) entre les 588 autres communes belges. La méthodologie sous-jacente à ce modèle est celle des choix discrets (TRAIN, 2003).

En ce qui concerne les modèles de mobilité quotidienne, on trouve d'abord le modèle de demande (2) qui est alimenté par deux entrées : le nombre de personnes actives issu du modèle de migration (1a) et le nombre d'emplois par commune tiré du modèle d'évolution (1b). Ces données constituent les marges de la matrice O/D qui sera calculée par un modèle gravitaire. L'étape suivante est un modèle de répartition modale (3). Son objectif est d'extraire de la matrice O/D originelle, qui concerne tous les modes, les déplacements réalisés en voiture. Cette procédure est nécessaire afin de fournir des données pertinentes au modèle d'accessibilité en véhicule privé (4) qui est une autre pierre essentielle de l'édifice. Ce dernier se base sur un équilibre de WARDROP pour l'affectation du trafic routier sur le réseau et fournit ainsi une matrice des temps de parcours entre communes. Ceux-ci sont enfin utilisés pour calculer des indices d'accessibilité (aux emplois et aux services) pour chacune des communes (KLEIN et alii, 2010). Ils serviront alors dans le modèle de localisation pour mesurer l'attraction de ces communes.

En vue d'employer cette chaîne de modèles dans une perspective de prospective, il est nécessaire de mettre à jour les valeurs prises en entrée du modèle de migration résidentielle afin de tenir compte des évolutions socio-démographiques et ce au niveau individuel puisqu'il s'agit là de l'échelle

prise en compte dans les différents modèles. C'est le rôle des modèles d'évolution (1b) qui utiliseront les populations synthétiques décrites dans la section précédente ainsi que des projections démographiques, etc.

### 3.3. *Modèle de propension à migrer*

Revenons-en à la première étape du modèle de mobilité résidentielle : le modèle de propension à migrer.

Sur base d'une revue de la littérature (p.ex. DEBRAND et alii, 2005 ou HENLEY, 1998), les variables explicatives ont été choisies parmi les données disponibles. Celles-ci sont de deux types : d'une part des « instantanés » du Registre National de la population à des 1er janvier successifs entre 2001 et 2005 et d'autre part le recensement (enquête socio-économique) d'octobre 2001 pour d'autres variables non accessibles via le Registre National. Il faut bien noter que l'utilisation de ces données désagrégées au niveau individuel est soumise à des règles très strictes et demande l'accord de la Commission de Protection de la Vie Privée (dans notre cas, les données n'étaient accessibles que via le GÉDAP qui y a accès pour les projections démographiques de la Belgique et uniquement dans leurs infrastructures). Rappelons que le niveau de granularité spatiale des données est celui de la commune. Ainsi, nous ne pouvons observer un déménagement que s'il implique un changement de commune entre deux 1er janvier successifs.

Dans ces ensembles de données, nous avons exploré le caractère explicatif de la propension à migrer de caractéristiques

individuelles telles que l'âge, le sexe, la nationalité, le type de ménage, le nombre d'individus dans le ménage, la position (lien avec le chef de ménage) dans le ménage, le niveau d'éducation ou le statut d'activité ; relatives au logement telles que le type de celui-ci (appartement, maison, ...) ou le fait d'en être propriétaire ou locataire.

De par la nature des données disponibles via les instantanés du Registre National, ce sont aussi des transitions entre états qui ont été retenues comme variables explicatives ; ainsi, par exemple, le fait de passer d'un 1er janvier au suivant d'un type de ménage « célibataire » au type « couple » peut avoir un impact sur la propension à migrer.

Avec ces variables, différents modèles ont été testés et calibrés. Au travers de ceux-ci, l'objectif était de tenir compte d'une certaine dynamique dans la propension à migrer. Suivant les modèles, on visait ainsi à tenir compte des effets de modifications d'état intervenues un an ou deux ans avant le déménagement ou de transitions intervenant dans l'année suivant celui-ci (p.ex. une naissance qui aurait été « préparée » par un changement résidentiel).

Comme nous disposions, via le Registre National, de données sur les dix millions d'individus résidant en Belgique mais que cette masse de données

nous semblait trop importante pour un travail de calibration efficace (notamment à cause de problèmes de temps de calcul et de stockage), nous avons effectué celui-ci sur un échantillon reprenant un cinquième de la population. Mais, comme la migration résidentielle (impliquant, rappelons-le, un changement de commune) reste un phénomène relativement rare (environ 4 % de la population s'y trouvent confrontés entre deux 1<sup>er</sup> janvier successifs), nous avons suréchantillonné le phénomène, en considérant dans l'échantillon 15 % de migrants, afin d'améliorer la qualité des coefficients de la régression (ALLISON, 1999). Cet échantillon a enfin été divisé en deux parties, l'une (70 %) servant à la calibration du modèle, l'autre à sa validation.

La qualité de chacun des modèles développés (pour le détail de ceux-ci, voir BAHRI et alii, 2009) a été appréciée au vu de différents diagnostics statistiques tels que le critère AIC par exemple. La Table 1 nous montre ainsi les tests sur le ratio de vraisemblance pour les cinq modèles testés.

*Table 1 : Tests statistiques appliqués aux différents modèles de propension à migrer*

	Tests sur les ratios de vraisemblance			
	Nombre de variables explic.	Khi-2	Ddl	Pr > Khi-2
Modèle 1	12	502377,9	215	< 0,0001
Modèle 2	14	505456,7	301	< 0,0001
Modèle 3	14	511072,5	219	< 0,0001
Modèle 4	17	511247,3	306	< 0,0001
Modèle 5	20	506794,7	394	< 0,0001

Ces tests nous montrent que chaque modèle a un bon pouvoir de prédiction. Leur valeur discriminante est également appréciable puisque, par exemple, la statistique c, qui est une approximation de la surface sous la courbe ROC, a une valeur proche de 0,9 alors que l'on considère (HOSMER, LEMESHOW, 2000) qu'une surface entre 0,8 et 0,9 est le signe qu'un modèle a un excellent pouvoir discriminant.

Par souci d'efficacité et d'économie (le bon vieux principe du rasoir d'Ockham), nous avons finalement retenu le plus simple de ces modèles, appelé Modèle 1 dans la Table 1, puisque les tests statistiques ne permettaient pas une réelle discrimination basée sur ceux-ci. Dans celui-ci, les variables explicatives retenues étaient :

- . âge en l'année A
- . sexe
- . nationalité en l'année A
- . type et taille (variable composite) du ménage en l'année A
- . évolution du type de ménage entre l'année A-1 et l'année A
- . évolution du type de ménage entre l'année A et l'année A+1

- . position par rapport au chef de ménage en l'année A
- . évolution de la position par rapport au chef de ménage entre l'année A-1 et l'année A
- . évolution de la position par rapport au chef de ménage entre l'année A et l'année A+1
- . déménagement entre l'année A-1 et l'année A
- . niveau d'éducation (plus haut diplôme obtenu)
- . statut d'occupation du logement (propriétaire ou locataire)

et la variable à expliquer : le déménagement (ou non) entre l'année A et l'année A+1.

De la calibration de ce modèle, il ressort que les facteurs les plus explicatifs de la propension à migrer sont, par ordre décroissant :

- . l'évolution du type et de la taille du ménage entre les années A et A+1 ;
- . le type de « possession » (propriétaire ou locataire) de son logement ;
- . l'évolution du lien de l'individu avec le chef de famille entre les années A et A+1 (c'est, par exemple, le cas d'un enfant qui devient chef de ménage ou conjoint c.-à-d. le cas du passage d'un rôle dans un ménage à un autre rôle dans un autre ménage) ;
- . l'âge (en considérant six classes d'âges) en l'année A.

Comparée à la situation de référence (pas de modification dans la structure du ménage entre deux années consécutives), n'importe quelle modification de la structure du ménage est corrélée avec une probabilité plus importante de déménager. La valeur de cette augmentation varie suivant les changements ; ainsi une personne dont la situation familiale était « couple marié avec enfants » mais devient « couple non marié avec ou sans enfant », « famille monoparentale » ou « isolé » a 20 à 30 fois plus de chances de changer de commune de résidence que quelqu'un dont la structure de ménage n'a pas changé. Dans le même ordre d'idées, une modification du lien avec le chef de famille, par exemple passer de « enfant » à « chef de ménage » ou « conjoint », entraîne une probabilité de déménagement entre 11 et 16 fois plus importante que celle d'une personne dont le statut n'a pas été modifié.

Seconde en ordre d'importance dans les facteurs prédictifs du déménagement, on trouve la variable relative au type de « possession » de son logement : un locataire dans le secteur privé aura 4 fois plus de chances de déménager qu'un propriétaire.

L'âge enfin est également un des quatre principaux facteurs mis en avant. Les personnes les moins enclines à changer de commune de résidence sont les 75 ans et plus alors que les plus mobiles sont les plus jeunes (moins de 18 ans), cette tendance à déménager diminuant avec l'âge (toutes choses étant égales).

Tous ces résultats démontrent que la propension à migrer est fortement liée

au cycle de vie des individus et plus spécifiquement à leurs trajectoires familiales. En résumé, les transitions amenant au déménagement sont les situations de ruptures, les (re-)compositions de nouvelles cellules familiales et le départ de la maison des parents. On peut aussi observer que les situations les plus stables concernent les personnes vivant mariées en couple (avec ou sans enfant) ; cette situation est souvent associée à un statut de propriétaire qui est un autre facteur de stabilité. On peut donc clairement supposer que les évolutions probables des cellules familiales marquées par une augmentation de ménages moins stables (cohabitants, familles monoparentales...) vont générer une plus grande propension à changer de commune de résidence dans les prochaines années.

### 3.4. MODÈLE DE LOCALISATION

Une fois la décision de déménager prise, il faut encore choisir quelle sera la nouvelle commune de résidence. C'est ce choix qui a été modélisé au travers du modèle de localisation.

Dans celui-ci, contrairement au cas du modèle de propension à migrer, il faut également tenir compte des caractéristiques des communes dans les variables explicatives du choix d'une nouvelle localité de résidence.

Une analyse préliminaire a donc été entreprise pour mettre en relief les caractéristiques communales les plus discriminantes. Dans un premier temps, sur base d'un échantillon de 100 000 personnes ayant déménagé entre les 1er janvier 2001 et 2002, il ressort que les mobilités résidentielles se font souvent à courte distance : environ 50 % de celles-ci se font vers une commune située à moins de 10 km de la commune de départ et seulement 10 % impliquent une distance de 50 km ou plus. Ensuite, les variables disponibles pour la modélisation ont été testées, tout à tour, dans un modèle avec une seule variable explicative via le logiciel BIOGEME (BIERLAIRE, 2003). Cela nous a permis de voir que, lorsqu'elles sont prises individuellement, la variable la moins explicative est l'indicateur du prix de l'immobilier (moyenne des prix de vente des maisons et appartements dans la commune) alors que celle qui s'est avérée la plus explicative est la distance entre la commune d'origine et la commune d'arrivée après déménagement.

L'étape suivante a consisté à mettre au point le modèle de choix discrets qui pourraient rendre compte du choix d'une nouvelle commune de résidence lors d'un déménagement. Une des difficultés à ce niveau vient du fait que, dans pareil choix, 588 possibilités s'offrent au décideur (puisque la Belgique compte 589 communes). Travailler avec un ensemble de choix de cette ampleur n'est pas courant dans la littérature et nous avons donc finalement opté pour un modèle du type Logit emboîté (TRAIN, 2003) où une typologie en 4 types (agglomération urbaine, banlieue, zone migratoire, rural) des communes belges (VAN DER HAEGEN et alii, 1996) a été utilisée. Des catégori-

sations plus « sophistiquées » croisant cette typologie avec d'autres critères géographiques (tels que l'appartenance à l'une des 3 régions -Flandre, Wallonie, Bruxelles- de Belgique) ont conduit par contre à des modèles avec bien davantage de paramètres non significatifs mais surtout avec un pouvoir explicatif moindre (log-vraisemblance et Rho carré ajusté de moins bonne qualité).

La Table 2 nous donne les résultats de la calibration de ce modèle pour toutes les variables significatives.

Ces résultats nous apprennent que les facteurs qui ont un poids dans le processus de décision des individus concernant leur nouvelle commune de résidence sont:

- . la distance à la commune de résidence précédente ; l'utilité (l'attractivité) d'une commune décroît au fur et à mesure qu'on s'éloigne de la commune de départ ; c'est cette variable qui est la plus explicative du choix de la nouvelle commune de résidence ;
- . l'accessibilité aux emplois (calculée comme le nombre d'emplois accessibles en moins de 30 min en voiture) ; l'accessibilité aux services n'a pas été retenue car elle s'est avérée très corrélée ( $r=0,89$ ) avec celle aux emplois ; cette accessibilité aux emplois donne de meilleurs résultats dans l'expression de l'utilité lorsqu'elle est déclinée sous forme de deux paramètres : l'un pour les personnes actives (comprises comme les personnes entre 19 et 64 ans), l'autre pour les personnes inactives ;
- . les quatre composants représentant les conditions de vie dans la commune (respectivement relativement à l'environnement, aux immeubles, aux conditions socio-économiques et aux services) sont eux aussi des facteurs explicatifs significatifs ; ces facteurs sont en fait des indicateurs composites basés sur différentes caractéristiques des communes et sur les appréciations (suggestives) que les répondants au recensement de 2001 devaient fournir sur les conditions de vie dans leur commune ; ils ont été calculés par le GÉDAP ;
- . il en va de même pour le prix moyen de l'immobilier dans la commune.

Les caractéristiques des individus (qui doivent faire le choix d'une nouvelle commune de résidence) ont été introduites dans le modèle comme des variables de contextualisation. En effet, si l'on considère uniquement les caractéristiques des individus décideurs comme variable explicative, il est clair que cela conduirait à des termes identiques dans l'utilité de chacune des possibilités (à savoir les 588 communes). Comme le choix de localisation repose sur des différences d'utilité entre communes, ces termes seraient non explicatifs et mèneraient à des modèles non identifiables. Nous avons donc plutôt introduit dans l'utilité de chaque commune, pour ces caractéristiques individuelles, des termes pour chaque catégorie (p.ex. pour chaque classe d'âges)



correspondant au produit d'une indicatrice, variable binaire reprenant si un individu appartient ou non à cette catégorie, et de la proportion représentée par cette catégorie dans la population de la commune (p.ex., pour un jeune, le pourcentage de jeunes dans la commune).

*Table 2 : Paramètres du modèle de localisation*

Paramètre	Valeur	Std err	t-test	p-value
Indicateur d'accessibilité aux emplois pour les personnes actives	0.0417	0.0208	2.01	0.04
Indicateur d'accessibilité aux emplois pour les personnes non actives	-0.175	0.0421	-4.16	0.00
Distance entre commune de départ et nouvelle commune	-0.0565	0.000433	-130.38	0.00
Indicateur de conditions de vie (aspects environnementaux)	1.08	0.201	5.37	0.00
Indicateur de conditions de vie (aspects logement)	-3.30	0.247	-13.38	0.00
Indicateur de conditions de vie (aspects socio-économiques)	4.72	0.195	24.27	0.00
Indicateur de conditions de vie (aspects services)	-3.45	0.285	-12.11	0.00
Indicateur du prix de l'immobilier	-0.373	0.163	-2.28	0.02
Population de la nouvelle commune	0.271	0.00987	27.43	0.00
Âge : 0 à 18	-17.7	6.49	-2.73	0.01
Âge : 19 à 29	11.7	1.05	11.15	0.00
Âge : 30 à 44	11.4	1.26	9.03	0.00
Âge : 45 à 54	5.77	2.48	2.33	0.02
Âge : 55 à 64	31.9	2.44	13.07	0.00
Âge : 65 à 74	10.7	2.75	3.88	0.00
Âge : 75 et plus	-2.20	3.57	-0.62	0.54
Éducation : autre	1.09	0.577	1.88	0.06
Éducation : secondaire inférieur	3.18	1.20	2.66	0.01
Éducation : secondaire supérieur	4.28	0.852	5.02	0.00
Éducation : enseignement supérieur	4.05	0.581	6.98	0.00
Type de ménage : autre	3.20	2.59	1.24	0.22
Type de ménage : couple (marié ou non) avec enfant(s)	1.28	0.332	3.87	0.00
Type de ménage : couple (marié ou non) sans enfant	5.95	0.637	9.34	0.00
Type de ménage : célibataire	12.0	0.472	25.38	0.00
Type de ménage : famille mono-parentale	7.78	0.716	10.87	0.00
Nationalité : autre	3.55	0.818	4.34	0.00
Nationalité : belge	4.07	0.324	12.54	0.00
Nationalité : pays frontaliers	1.33	1.52	0.87	0.38
Nationalité : reste de l'Union Européenne	6.04	1.83	3.31	0.00

Ces variables de contextualisation sont significatives et indiquent que les individus tendent à s'installer dans des communes où ils peuvent retrouver leurs semblables (c.-à-d. des individus ayant les mêmes caractéristiques). Tant les classes d'âges que les types de ménages, les nationalités ou les niveaux d'éducation font apparaître cette tendance prouvant l'envie des individus de vivre avec des personnes des mêmes groupes socio-économiques.

#### 4. CONCLUSION

L'outil que constitue la chaîne de modèles développée dans le cadre de MOBLOC peut maintenant être valorisé dans des études prospectives où elle servira à estimer les migrations (mobilité résidentielle) de la population belge en fonction de son évolution démographique, de scénarios liés tant à l'aménagement du territoire qu'à la mobilité quotidienne (et son impact sur l'accessibilité). La micro-simulation des changements intervenant chez les individus (p.ex. acquisition d'un nouveau diplôme) ou dans les ménages (p.ex. séparation ou naissance) réalisée grâce à l'outil des populations synthétiques et des techniques permettant leurs évolutions est, dans cette perspective, indispensable pour simuler les tendances dans les migrations résidentielles. En effet, nous avons vu que ce sont de tels événements qui sont les facteurs les plus explicatifs de la propension à migrer.

En outre, les populations synthétiques et les méthodes permettant de faire évoluer celles-ci seront également nécessaires dans ce contexte pour réaliser des projections futures qui soient pertinentes à un niveau de désagrégation spatiale fin en vue de mesurer des impacts de mesures à un niveau proche des collectivités locales et d'être à même de différencier les effets de politique en fonction des territoires caractérisés par leur aménagement propre.

Ainsi en appliquant dans chaque période du futur considéré, les modèles de mobilité résidentielle sur des populations qui ne seront pas des clones de la population actuelle mais qui auront évolué, d'un point de vue sociodémographique, parallèlement à leur distribution géographique redessinée par les déménagements, il sera possible de mieux simuler les évolutions à long terme qui marqueront l'aménagement du territoire comme la mobilité (quotidienne).

En outre, les modèles d'accessibilité, également développés dans le cadre de MOBLOC mais non détaillés ici, permettront également d'évaluer les interactions entre mobilité quotidienne et mobilité résidentielle sur des horizons du moyen et du long terme. Ils permettront aussi de modéliser certaines politiques de transport, d'en mesurer l'impact sur l'accessibilité et, à travers cela et grâce à la chaîne de modèles développés, sur la mobilité résidentielle. Par exemple, il sera ainsi possible d'estimer comment des mesures visant à « pénaliser » le transport routier pourront jouer un rôle dans le ralentissement de l'étalement urbain.

Enfin, si les méthodologies développées l'ont été dans le cadre de la Belgique, il n'en demeure pas moins vrai qu'elles sont relativement générales et peuvent donc être adaptées à d'autres contextes, à d'autres pays moyennant la disposition des données nécessaires voire une adaptation des modèles aux données effectivement disponibles. Cela permettrait de voir dans quelle mesure les déterminants de la mobilité résidentielle sont les mêmes partout ou si l'impact de l'accessibilité sur le choix d'une nouvelle commune de résidence reste invariant quelle que soit la région.

#### REMERCIEMENTS

Il nous faut tout d'abord remercier la Politique Scientifique Fédérale Belge (BELSPO) qui a financé le projet MOBLOC. Nos remerciements vont aussi aux équipes de recherche qui ont collaboré avec nous sur ce projet : celle du GéDAP (UCL) sous la direction de Thierry EGGERICKX et celle de Géode (CEPS-INSTEAD) sous la direction de Philippe GERBER. Nous tenons également à remercier l'Académie Louvain qui a soutenu le projet DIDAM dans lequel les méthodes de construction de populations synthétiques ont été développées. Enfin, nous remercions les deux lecteurs qui nous ont fournis des remarques pertinentes et constructives.

#### BIBLIOGRAPHIE

- ALLISON P. (1999) **Logistic Regression Using SAS Theory and Application**. SAS Press.
- BAHRI A., CARPENTIER S., CORNELIS E., EGGERICKX T., GERBER Ph., KLEIN S., PAULY X., TOINT Ph., WALLE F. (2009) **Mobilities and long term location choices in Belgium "MOBLOC" Final Report Phase 1**. Brussels, Belgian Science Policy (Research Programme Science for a Sustainable Development), 47 p.
- BARTHELEMY J., CORNELIS E., JOURQUIN B., PIOTTE J., TOINT Ph. (2010) Towards microsimulation of passenger and freight transport competition: Advances in synthetic population generation and simulation of the behaviour of freight actors. **Proceedings of the 12th WCTR**, Lisboa, 10-15 juillet, 21 p.
- BARTHELEMY J., TOINT Ph. (2012) Synthetic population generation without a sample. **Transportation Science**, à paraître.
- BECKMAN R.J., BAGGERLY K.A., MCKAY M.D. (1996) Creating synthetic baseline populations. **Transportation Research Part A**, Vol. 30, n° 6, pp. 415-429.
- BIERLAIRE M. (2003) BIOGEME: A free package for the estimation of discrete choice models. **Proceedings of the 3rd Swiss Transportation Research Conference**, Ascona.

- CARUSO G., PEETERS D., CAVAILHES J., ROUNSEVELL M. (2007) Spatial configurations in a periurban city. A cellular automata-based microeconomic model. **Regional Science and Urban Economics**, Vol. 37, n° 5, pp.542-567.
- CAVAILHES J., PEETERS D., SEKERIS E., THISSE J.-F. (2004) The periurban city: why to live between the suburbs and the countryside, **Regional Science and Urban Economics**, Vol. 34, n° 6, pp. 681-703.
- CHOOCHARUKUL K., VAN H.T., FUJII S. (2008) Psychological effects of travel behavior on preference of residential location choice. **Transportation Research Part A: Policy and Practice**, Vol. 42, n° 1, pp. 116-124.
- CIRILLO C., CORNELIS E., TOINT Ph. (2012) A Model of Weekly Labor Participation for a Belgian Synthetic Population. **Networks and Spatial Economics**, Vol. 12, n° 1, pp. 59-73.
- CORNELIS E., LEGRAIN L., TOINT Ph. (2005) Synthetic populations: a tool for estimating travel demand. In B. JOURQUIN (ed.) **BIVÉC-GIBET Transport Research Day 2005**, Vol. 1, Brussels University Press, pp. 217-235.
- DEBRAND T., TAFFIN C. (2005) Les facteurs structurels et conjoncturels de la mobilité résidentielle depuis 20 ans. **Économie et statistique**, n° 381-382, pp. 125-146.
- DEMING W.E., STEPHAN F.F. (1940) A least squares adjustment of a sampled frequency table when the expected marginal totals are known. **Annals of Mathematical Statistics**, 11, pp. 428-444.
- DUJARDIN C., SELOD H., THOMAS I. (2008) Residential segregation and unemployment: the case of Brussels. **Urban Studies**, Vol. 45, n° 1, pp. 89-113.
- FRICK M., AXHAUSEN K. (2004) Generating synthetic populations using IPF and Monte Carlo techniques: some new results. **Technical Report, conference paper STRC 2004**.
- GUO Y., BHAT C.R. (2007) Population synthesis for the microsimulating travel behavior. **Transportation Research Record: Journal of the Transportation Research Board**, n° 2014, pp. 92-101.
- HENLEY A. (1998) Residential mobility, housing equity and the labour market. **The Economic Journal**, 108, pp. 27-414.
- HOSMER D.W., LEMESHOW S. (2000) **Applied Logistic Regression**. Wiley, 373 p.
- HUANG Z., WILLIAMSON P. (2002) **A comparison of synthetic reconstruction and combinatorial optimization approaches to the creation of small-area microdata**. University of Liverpool, Department of Geography, Working paper.

- KLEIN S., OMRANI H., CARPENTIER S., GERBER Ph., VANDENBULCKE G. (2010) Validation d'un modèle d'accessibilité par recouplement de données multi-sources. Application aux communes de Belgique. **2ème Journée de Recherche Mobilité, Transport et Logistique (MTL 2010)**, Lyon, 23 juin, 16 p.
- LINDBERG E., HARTIG T., GARVILL J., GÄRLING T. (1992) Residential-Location Preferences Across The Life-Span. **Journal of Environmental Psychology**, Vol. 12, n° 2, pp. 187-198.
- NG C.F. (2008) Commuting distances in a household location choice model with amenities. **Journal of Urban Economics**, Vol. 63, n° 1, pp. 116-129.
- NIJKAMP P., VAN WISSEN L., RIMA A. (1993) A Household Life-Cycle Model For Residential Relocation Behavior Socio-Economic. **Planning Sciences**, Vol. 27, n° 1, pp. 35-53.
- ORCUTT G. H. (1957) A New Type of Socio-Economic System. **The Review of Economics and Statistics**, Vol. 39, n° 2, pp. 116-123.
- ORCUTT G., CALDWELL S., WERTHEIMER R. (1976) **Policy exploration through microanalytic simulation**. Urban Institute, Governance in Europe Series.
- ROSSI P.H. (1955) **Why Families Move: A Study in the Social Psychology of Urban Residential Mobility**. Free Press, 220p.
- SPIELAUER M. (2009) **Qu'est-ce qu'une microsimulation dynamique en sciences sociales ?** Statistics Canada, Modeling Division.
- TOINT Ph., EGGERICKX T., GERBER Ph., CARPENTIER S., CORNELIS E., DAL L., KLEIN S., PAULY X., WALLE F. (2011) **Mobilities and long term location choices in Belgium "MOBLOC"-Final Report**. Brussels, Belgian Science Policy (Research Programme Science for a Sustainable Development), 81 p.
- TRAIN K.E. (2003) **Discrete Choice Methods with Simulation**. Cambridge University Press, 334 p.
- TURNER M. (2005) Landscape preferences and patterns of residential development. **Journal of Urban Economics**, Vol. 57, n° 1, pp. 19-54.
- VAN DER HAEGEN H., VAN HECKE E., JUCHTMANS G. (1996) Les régions urbaines belges en 1991. **Études Statistiques**, n° 104.
- VOAS D., WILLIAMSON P. (2001) An evaluating goodness-of-fit measures for synthetic microdata. **Geographical and Environmental Modeling**, Vol. 5, n° 2, pp. 177-200.
- WILSON A.G., POWNALL C.E. (1976) A new representation of the urban system for modeling and for the study of microlevel interdependence. **Area**, 8, pp. 246-254.
- ZEGRAS P.C., SRINIVASAN S. (2007) Household income, travel behavior, location, and accessibility-Sketches from two different developing contexts. **Transportation Research Record**, n° 2038, pp. 128-138.